GAP Test

Ghosh, Anarta; Petkov, Nicolai

GAP Test : A Cognitive Evaluation Procedure for Shape Descriptors

Anarta Ghosh and Nicolai Petkov
Institute of Mathematics and Computing Science,
University of Groningen, P.O. Box 800, 9700 AV,
Groningen, The Netherlands. Email: anarta@cs.rug.nl, petkov@cs.rug.nl

Abstract

With inspiration from psychophysical researches of the human visual system we propose a novel method for performance evaluation of contour based shape recognition algorithms. We use complete contour representations of objects as a training set. Incomplete contour representations of the same objects are used as a test set and the recognition performance of two shape based methods is investigated. The amount of incompleteness in test cases is quantified using the percentage of contour pixels retained. The performances of the methods are reported using the recognition rate as a function of the degree of incompleteness. We consider three types of incomplete contour representations, viz. segment-wise deletion, occlusion and random pixel depletion. The methods compared in this framework use shape context and distance multiset as local shape descriptors. Qualitatively, both methods mimic human visual perception in the sense that they perform best in the case of random depletion and worst in the case of occluded contours. Quantitatively, the distance multiset method performs better than the shape context method in this test framework.

1. Introduction

If we look at the objects in the Figure 1 we can instantly recognize them as birds, even though 80% of the contour points have been removed (randomly) in the left image, the right half of the contour is not visible in the middle image, and 50% of the contour is removed (segment-wise) in the right image. This ability of human beings to recognize objects with incomplete contour representations was studied by the psychologist E. S. Gollin [4]. His objective was to investigate the performance of humans in recognizing objects with incomplete representations as a function of developmental characteristics, such as mental and chronological age and intelligence quotient. The subjects of his experiments were children of different age groups and a group of adults. Gollin used sets of contour images with different degree of incompleteness (Figure 2) and addressed the following questions: (1) In order to be recognized by humans, how complete the representations of common objects need to be? (2) How does training affect the recognition performance in case of incomplete representations?

(a) (b) (c)

Figure 1. A bird can be recognized even though its contour is incompletely represented.

With inspiration from Gollin's work we propose a novel attribute, viz. *robustness to incomplete contour representations*, that any contour based object recognition system/algorithm should have. The objective of this study is to investigate the performance of recognition systems/algorithms in an idealized situation where: (a) complete contour representations of the objects to be recognized form the reference (training) set or "memory" of the system/algorithm, (b) incomplete contour representations of the same objects are derived from the afore mentioned complete representations, (c) the performance of the system/algorithm in recognizing the objects from these incomplete representations is evaluated. The main reason behind choosing such an ideal situation is the rational logic that in order to perform well in a real world scenario (natural images) any recognition system should first perform well in such ideal (simple) situations.

In the context of processing visual information using computers this aspect of recognition of objects with incomplete contours is also very important. Let us consider a natu-

Proceedings of the Fourth International Conference on Hybrid Intelligent Systems (HIS'04)
0-7695-2291-2/04 $ 20.00 IEEE

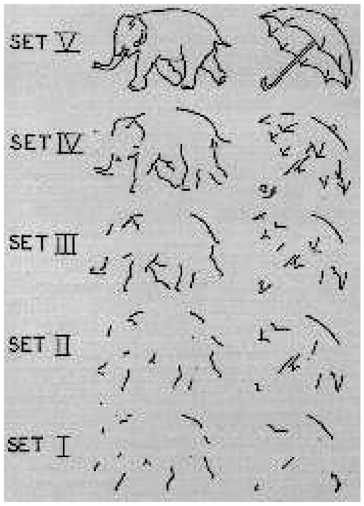

Figure 2. Example of image sets used in Gollin's original test [4].

(a) (b) (c)

Figure 3. (a) Image of a rhinoceros in it's natural habitat. (b) The result of edge detection with a bank of Gabor energy filters. (c) The result of contour detection by a bank of Gabor energy filters augmented with a biologically motivated surround suppression of texture edges.

ral image and two edge images, obtained from it (Figure 3). The middle image was obtained by applying a bank of Gabor energy filters. It contains the contour of the object of interest, viz. a rhinoceros, but it also contains a large number of texture edges that are not related in any way to the shape of the rhinoceros. These texture edges will have a devastating effect on the performance of all currently known contour based shape recognition methods. Advanced contour detection methods based on surround suppression succeed in separating the essential object contours from the texture edges, as illustrated by the right-most image in Figure 3, but at the same time these methods have a certain negative side effect of depleting the contours of the objects of interest. Hence the robustness of shape recognition methods to contour depletion is an issue of practical importance.

We consider three types of incomplete contour representations of objects and according to the way they are constructed we name the studies as follows: (1) *segment-wise deletion test,* (2) *occlusion test,* (3) *depletion test.* In short we call these tests GAP tests.

The choice of the shape recognition methods we study is limited by the condition that they use contour images as inputs. Methods which use other type of information fall outside the scope of this study without modification. For instance, Gavrila [3] proposes a method based on the distance transform, in which every point of a solid binary object is characterized by its distance to the object's border. In our study objects are represented by their contour points only, and hence the distance transform is not informative. Latecki and Lakämper's polygonal shape descriptor [7] inherently assumes that an object is represented by a closed curve, and therefore this method cannot be applied to objects represented by incomplete contours. For further references to shape analysis and object recognition methods see e.g. [5, 9, 13].

In this paper we study the *shape context* method described in [1] and the *distance multiset* method described in [5] with respect to their robustness to contour deletions of different types. In Section 2 we briefly describe these methods. In Section 3 we present the experimental design and the achieved results. A discussion and conclusions are given in Section 4.

2. Contour Based Shape Recognition Methods

In both the shape context method and the distance multiset method the recognition of objects is done by computing dissimilarity between the contour representations of two objects by using a point correspondence paradigm. The point correspondences are found using shape descriptors associated with the points.

2.1. Shape Context

A shape descriptor, called the *shape context* [1], of a point p, belonging to the contour of an object is a two-dimensional histogram in a log-polar coordinate system that gives the distribution of contour points in the surroundings of p. Let an object $\mathcal{O}$ be represented by a set of contour points, $\mathcal{O} \equiv \{p_1 \dots p_N\}$. Formally, the authors of this method define the shape context of a point $p \in \mathcal{O}$ as a vector in the following way,

$$H^{\mathcal{O}}_{K,p} = \{h_p(k) : k = 1 \dots K\} \qquad (1)$$

where,

$$h_p(k) = card\{q \neq p | q \in \mathcal{O}, (q-p) \in bin(k)\} \quad (2)$$

and K is the total number of histogram bins. The bins are constructed by dividing the image plane into K partitions of equal size (in a log-polar coordinate system) with p as the origin. In this study we use 5 intervals for the log distance, and 12 intervals for the polar angle, so K = 60. To improve performance, in this study we normalize $H^{\mathcal{O}}_{K,p}$ by the total number of contour pixels of the concerned object. The shape of the object is described using the set of shape contexts associated with all contour points in the following way:

$$S^{SC}_{\mathcal{O}} \equiv \{H^{\mathcal{O}}_{K,p} | p \in \mathcal{O}\}. \quad (3)$$

The cost of matching a point p_i belonging to the contour of an object $\mathcal{O}_1$, having M points, to a point q_j belonging to the contour of an object $\mathcal{O}_2$, having N points is defined as follows:

$$c^{SC}_{i,j} \equiv \frac{1}{2} \sum_{k=1}^{K} \frac{[h_{p_i}(k) - h_{q_j}(k)]^2}{[h_{p_i}(k) + h_{q_j}(k)]} \quad (4)$$

An $M \times N$ cost matrix of point-wise dissimilarities is constructed according to (4). Next we compute the dissimilarity between the shapes $S^{SC}_{\mathcal{O}_1}$ and $S^{SC}_{\mathcal{O}_2}$ of the objects in the following way:

$$d^{SC}(S^{SC}_{\mathcal{O}_1}, S^{SC}_{\mathcal{O}_2}) \equiv \sum_{i=1}^{M} min\{c^{SC}_{i,j} | j = 1, \dots, N\} \quad (5)$$

The authors of the shape context approach [1] use a different method to compute the dissimilarity of two shapes from the point-wise dissimilarity matrix. More specifically they use the Hungarian algorithm of bipartite graph matching to solve the optimal assignment problem. In our experiments we found that the above mentioned method gives sufficient results.

2.2. Distance Multiset

For a point p in the contour of an object $\mathcal{O}$, having N points, the *distance multiset* is formally defined as follows [5]:

$$D^{\mathcal{O}}_{N,p} = \{log(d_j(p)) | j = 1 \dots N-1\} \quad (6)$$

where $d_j(p)$ is the Euclidean distance between p and its j^{th} nearest neighbor in $\mathcal{O}$. In this approach the shape of an object $\mathcal{O} \equiv \{p_1 \dots p_N\}$ defined by set of points, is described by the set of distance multisets in the following way:

$$S^{DM}_{\mathcal{O}} \equiv \{D^{\mathcal{O}}_{N,p} | p \in \mathcal{O}\}. \quad (7)$$

Next, a cost of matching two distance multisets is defined as follows. Consider the sets/multisets,

$$X = \{x_i \in \mathbf{R} | i = 1, \dots, M\} \quad (8)$$

$$Y = \{y_i \in \mathbf{R} | i = 1, \dots, N\} \quad (9)$$

where $M \leq N$. Let A be the following $M \times N$ matrix of absolute values of pair-wise differences of elements of X and Y:

$$A_{i,j} = |x_i - y_j|, \; i = 1 \dots M, j = 1 \dots N. \quad (10)$$

Let π be a one-to-one mapping from the set $\{1, \dots, M\}$ to the set $\{1, \dots, N\}$ and let Π be the set of all such mappings. The mapping π defines an assignment of an unique element $y_{\pi(i)} \in Y$ to each element $x_i \in X$. The cost $c_\pi(X, Y)$ of a mapping/assignment $\pi \in \Pi$ is defined as follows:

$$c_\pi(X, Y) = \sum_{i=1}^{M} A_{i,\pi(i)} \quad (11)$$

Let c be the minimum of the costs of all such possible mappings :

$$c(X, Y) = min\{c_\pi(X, Y) | \pi \in \Pi\} \quad (12)$$

We now define the cost $c^{DM}_{i,j}$ of matching a point p_i in an object $\mathcal{O}_1$ represented by M contour points to a point q_j in an object $\mathcal{O}_2$ represented by N contour points, $M \leq N$, using the definition of c according to (6-12) in the following way:

$$c^{DM}_{i,j} \equiv c(D^{\mathcal{O}_1}_{N,p_i}, D^{\mathcal{O}_2}_{M,q_j}) \quad (13)$$

Note that $D^{\mathcal{O}_1}_{N,p_i}$ and $D^{\mathcal{O}_2}_{M,q_j}$ are sorted in ascending order by the definition of a distance multiset. To compute $c(D^{\mathcal{O}_1}_{N,p_i}, D^{\mathcal{O}_2}_{M,q_j})$ efficiently, we use the algorithm described in [12].

Similar to (5) the dissimilarity between the shapes $S^{DM}_{\mathcal{O}_1}$ and $S^{DM}_{\mathcal{O}_2}$ of the objects is defined as follows:

$$d^{DM}(S^{DM}_{\mathcal{O}_1}, S^{DM}_{\mathcal{O}_2}) \equiv \sum_{i=1}^{M} min\{c^{DM}_{i,j} | j = 1 \dots N\} \quad (14)$$

3. Experiments and Results

3.1. Dataset

As a data set we choose images obtained from the MPEG-7 database [8]. It contains 1400 images divided in 70 classes, each of 20 similar objects (eg. apple, bird, bat, etc). We choose one object from each class (Figure 4, row 1) and extract the contours of the object using Gabor filters [6]. The resulting 70 contour images are rescaled in such a way that the diameter (maximum distance between the contour pixels) is approximately the same for all objects, cf. row 2 of Figure 4. These 70 rescaled contour images are used as reference images in our experiments. The set of these images corresponds to the complete representations, set V of Figure

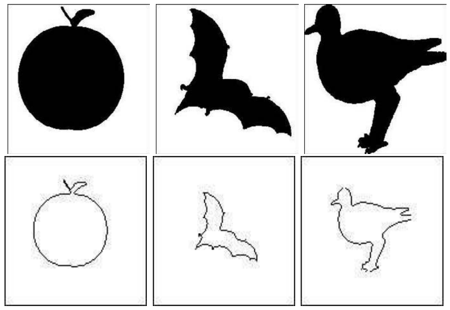

Figure 4. Row 1: Sample of the MPEG-7 database images used. Row 3: Corresponding rescaled contour images; these images are considered as complete representations that comprise the memory of the recognition system.

2, used in Gollin's original study and form the "memory" of the recognition system.

For the *segment-wise deletion test* incomplete representations (Figure 5 row 1) are constructed by randomly removing continuous segments of the contours and retaining a given percentage of contour pixels from the above mentioned complete contour representations.

For the *occlusion test* incomplete representations are created by removing a given percentage of consecutive contour pixels starting from the leftmost (Figure 5 row 2) or the rightmost pixel of an object.

For the *depletion test* the incomplete representations (Figure 5 row 3) are obtained by randomly removing a given percentage of pixels from the contours of the complete contour representations.

In our experiments the percentages of retained pixels are chosen in the following way: 2% to 4% in steps of 1%, 5% to 85% in steps of 5%, and 100% for the depletion tests, 5% to 85% in steps of 5%, and 100% for the segment-wise deletion test, and for the occlusion test. For each type (segment-wise deletion, occlusion and depletion) and degree of contour image degradation we create 70 test images. All complete contour images and incomplete contour images obtained with different types and percentages of incompleteness are available in the web-site *www.cs.rug.nl/∼petkov*.

3.2. Method.

An incomplete representation (segment-wise deleted or depleted or occluded contour image) obtained from one of the 70 reference images is compared with all 70 reference images and a decision is taken about which reference image the degraded image is most similar to (nearest neighbor search). The comparison is based on a shape dissimilarity computed using a given shape comparison method, described in Section 2. If the nearest neighbor is the reference image from which the degraded image was obtained, the recognition is considered correct, otherwise incorrect. If the nearest neighbor is found to be not unique then the recognition is also considered incorrect. For each of the three tests (segment-wise deletion, occlusion, depletion) and for each degree of contour image degradation, corresponding 70 test images are compared with each of the 70 reference images and the percentage of correct recognition is determined. The percentage of correct recognition is observed as a function of the percentage of retained contour pixels. In the case of occlusion test the percentage of correct recognition is calculated by averaging the correct recognition rates with left and right occluded images for a given percentage of retained contour.

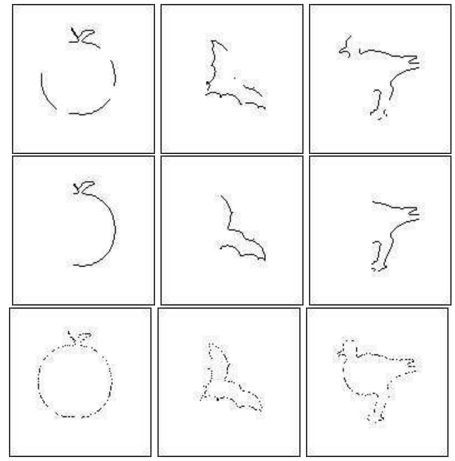

Figure 5. Incomplete Representations: Row 1: Segment-wise deleted contour representations of objects (they correspond to the incomplete representations of Gollin's original study, set I to IV of Figure 2). Row 2: Left occluded contour representations of objects Row 3: Depleted contour representations of objects.

3.3. Results

Figure 6, 7, 8 show the results of our experiments.

In the case of the segment-wise deletion test (Figure 6) and the occlusion test (Figure 7), the performance of distance multiset method is appreciably better than that of the shape context method for any percentage of retained contour pixels. From the results of the depletion test (Figure 8) we see that both the shape context method and the distance multiset method perform very well in recognizing objects

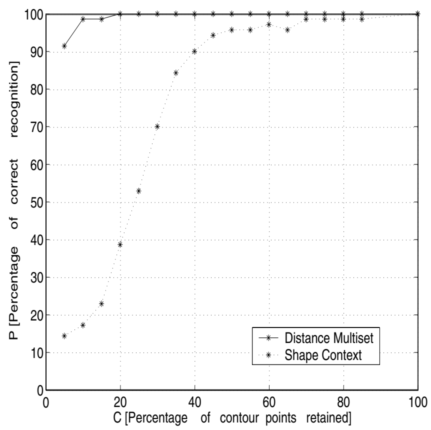

Figure 6. Results of the segment-wise deletion test.

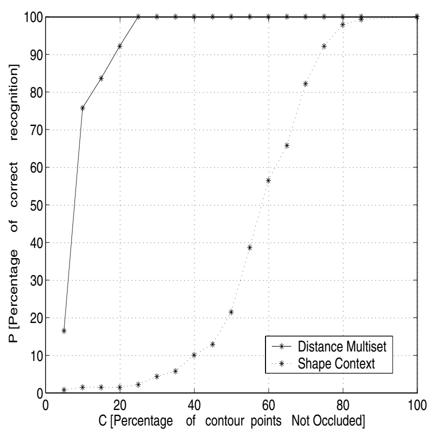

Figure 7. Results of the occlusion test.

with depleted contour representations, if more than 40% and 5%, respectively, of the contour points are retained. The distance multiset method outperforms the shape context method when the degree of depletion is very high, i.e. very low percentage (less than 40%) of the pixels are retained.

4. Discussion, Summary and Conclusion

Related Work Object recognition methods that employ shape descriptors have been evaluated and compared using various characteristics like invariance, uniqueness and stability [11]. Marr and Nishihara [10] proposed three criteria for judging the effectiveness of a shape descriptor, viz. accessibility, scope and uniqueness, stability and sensitiv-

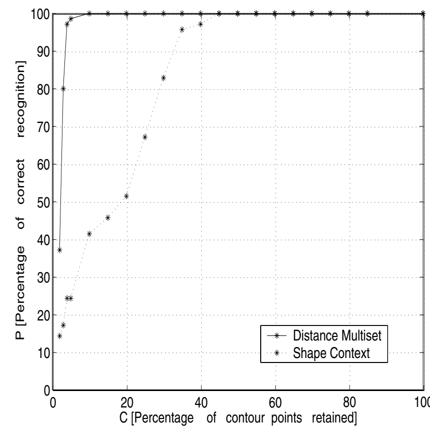

Figure 8. Results of the depletion test.

ity. Brady [2] puts forward a set of criteria for representation of shape, viz. rich local support, smooth extension and propagation. In the current work, motivated by characteristics of the human visual system [4], we propose an additional new criterion to compare and characterize contour based shape descriptors using their performance in recognizing objects with incomplete contours. We are not aware of any such comparison and characterization in the present literature.

Discussion To explain the high performance of the distance multiset method let us consider the sets $A, B, \subset \mathbf{R}^2$ such that, $B = \{f(a) : a \in A\}$, where $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ is defined as follows, $f(\mathbf{x}) = \mathbf{L}\mathbf{x} + \mathbf{t}, \forall \mathbf{x} \in \mathbf{R}^2$, $\mathbf{L}$ being a 2×2 orthogonal matrix and $\mathbf{t} \in \mathbf{R}^2$.

So if A is the set of contour points of an object $\mathcal{O}_1$ then B is the set of contour points of an object $\mathcal{O}_2$ that is derived from $\mathcal{O}_1$ through the transformation described above.

Let us recall an important result regarding functions in n dimensional Euclidean space : $\phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is an isometry (i.e an Euclidean distance preserving transformation) iff for any $\mathbf{X} \in \mathbf{R}^n$ $f(\mathbf{X}) = \mathbf{L}\mathbf{X} + \mathbf{t}$, $\mathbf{L}$ is an $n \times n$ orthogonal matrix and $\mathbf{t} \in \mathbf{R}^n$. We now formulate the following.
Lemma: If A, B are defined in the above mentioned way and

$$C \subset B,\ \mathrm{card}(C) \geq 2 \tag{15}$$

then,

$$d^{DM}(S_C^{DM}, S_A^{DM}) = 0 \tag{16}$$

where S_C^{DM} and S_A^{DM} are the shapes, described by distance multisets, corresponding to C and A, respectively.
Proof:
Claim 1 : The distance multisets of A and B are identical.

By definition f is an isometry, which implies that distance multisets of A and B are identical, that is, for every

$p \in A$ $\exists$ a $q \in B$, $q = f(p)$ such that $D^A_{N,p} = D^B_{N,q}$, assuming that card(A) = card(B) = N.

Claim 2 : $d^{DM}(S^{DM}_C, S^{DM}_B) = 0$.

The definition of distance multiset along with (13) and (15) implies that for every $p_i \in C$, $\exists$ a $q_j \in B$, such that $c^{DM}_{i,j} = 0$. Hence the minimum of every row of the cost-matrix of point-wise dissimilarities is 0, which implies by (14), $d^{DM}(S^{DM}_C, S^{DM}_B) = 0$.

Claim 2 and the invariance of distance multisets in claim 1, along with (13) and (14) imply that

$$d^{DM}(S^{DM}_C, S^{DM}_A) = 0 \; \Box.$$

The implication of the lemma is two-fold : (1) In the case of the distance multiset method, the recognition will be incorrect in this test framework only when the nearest neighbor of a test object in the reference set is not unique. (2) In our study A corresponds to the set of contour points of a reference object, f is identity transformation and C is the set of contour points corresponding to an incomplete representation. *Theoretically*, the distance multiset method should perform exactly the same way when f is not identity transformation.

Both methods perform worst in the occlusion test and best in depletion test, which conforms with the recognition performance of humans, as occluded contour images carry the least and depleted contour images carry the maximum shape information.

Summary and Conclusion. With inspiration from Gollin's work [4] we proposed a method for evaluation of contour based shape recognition algorithms. To summarize the test framework, we put forward the following procedure,

GAP test : (Step 1) Take a set of images of objects and extract contours. Rescale the contour images in such a way that the diameter of the objects are approximately same, say lie between 70 and 76 pixel units. (Step 2) Train the recognition system with these complete contour representations. (Step 3) Construct different sets of incomplete representations from the complete contour representations, quantifying the level of incompleteness using percentage of contour pixels retained. (Step 4) Evaluate the recognition performance of the system, by computing the recognition rate as a function of the percentage of contour pixels retained in the incomplete representations.

We created a test database and made it publicly available at *www.cs.rug.nl/~petkov/*.

We illustrated the test framework with two shape recognition methods based on the shape context and the distance multiset. *In this context we want to emphasize that the results presented in this paper are intended to explain the conceptual aspects of the GAP test framework and not to put forward a complete comparison of the methods.*

Our main conclusions about the research presented in this paper are as follows : The robustness of contour based shape recognition methods to incompleteness of contour representations is an important aspect of any contour based objects recognition system. The GAP test as defined and proposed in this paper is an adequate framework for assessing the above mentioned performance and can be used as a standard test procedure for any contour based object recognition system/algorithm.

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 24, No. 24*, pages 509–522, 2002.

[2] M. Brady. Criteria for representations of shape. *In J. Beck, B. Hope and A. Rosenfeld, editors, Human and Machine Vision, Academic Press*, pages 39–84, 1983.

[3] D. Gavrila. Multi-feature hierarchical template matching using distance transforms. In *International Conference on Pattern Recognition (ICPR '98)*, pages 439–444, Brisbane, Australia, 1998.

[4] E. Gollin. Developmental studies of visual recognition of incomplete objects. *Perceptual and Motor Skills. Vol. 11*, pages 289–298, 1960.

[5] C. Grigorescu and N. Petkov. Distance sets for shape filters and shape recognition. *IEEE Transactions on Image Processing, Vol. 12, No. 10*, pages 1274–1286, 2003.

[6] C. Grigorescu, N. Petkov, and M. Westenberg. Contour detection based on nonclassical receptive field inhibition. *IEEE Transactions on Image Processing, Vol. 12, July*, pages 729–739, 2003.

[7] L. Latecki and R. Lakämper. Shape similarity measure based on correspondence of visual parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 22*, pages 1185–1190, 2000.

[8] L. Latecki, R. Lakämper, and U. Eckhardt. Shape descriptors for non-rigid shapes with single closed contour. *In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 424–429, 1998.

[9] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition, Vol. 31, No. 8*, pages 983–1001, 1998.

[10] D. Marr and H. Nishihara. Representation and recognition of the spatial organization of three dimensional shapes. *In Proc. Roy. Soc. London, B., Vol. 200*, pages 269–294, 1978.

[11] F. Mokhtarian and A. Mackworth. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 8*, pages 789–805, 1992.

[12] N. Petkov. Algorithm for the cost of an optimal assignment of two sets of real numbers. *Technical report, 2003-9-07, Institute of Mathematics and Computing Science, University of Groningen*, 2003.

[13] R. Veltkamp and M. Hagedoorn. State of the art in shape matching. *Technical Report UU-CS-1999-27, Utrecht University*, 1999.